DOCUMENT RESUME

ED 438 321 TM 030 638

AUTHOR	Klemmer, Cynthia Davis
TITLE	Stepwise Descriptive or Predictive Discriminant Analysis: Don't Even Think about Using It!
PUB DATE	2000-01-29
NOTE	20p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Dallas, TX, January 27-29, 2000).
PUB TYPE	Information Analyses (070) -- Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Discriminant Analysis
IDENTIFIERS	Descriptive Discriminant Analysis; Predictive Discriminant Analysis; *Stepwise Regression

ABSTRACT

C. Huberty (1994) recently noted that "It is quite common to find the use of 'stepwise analyses' reported in empirically based journal articles." Stepwise methods are used (incorrectly) by some researchers either to select variables to retain for further analyses or to evaluate the relative importance of various variables. Of course, stepwise methods have been shown not to be useful for either purpose (B. Thompson, 1995, 1999). Various authors have presented scathing indictments of many of these applications (C. Huberty, 1989; P. Snyder, 1991; B. Thompson, 1989). This paper explains the three major problems with stepwise applications in the context of discriminant function analysis. The paper considers both descriptive discriminate analysis and predictive discriminant analysis, noting that these are two very different applications (C. Huberty, 1994, and B. Thompson, 1995). It is suggested that stepwise methods not be used in multi-variate statistics. (Contains 20 references.) (Author/SLD)

Stepwise Descriptive or Predictive Discriminant Analysis: Don't Even Think About Using It!

Cynthia Davis Klemmer

Texas A&M University

Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, January 29, 2000.

Abstract

Huberty (1994) recently noted that, "It is quite common to find the use of 'stepwise analyses' reported in empirically based journal articles" (p. 261). Stepwise methods are used (incorrectly) by some researchers either to select variables to retain for further analyses or to evaluate the relative importance of various variables. Of course, stepwise methods have been shown to not be useful for either purpose (Thompson, 1995, 1999). Indeed, various authors have presented scathing indictments of many of these applications (cf. Huberty, 1989; Snyder, 1991; Thompson, 1989). The present paper will explain the three major problems with stepwise applications, and will do so in the context of discriminant function analysis. The paper will consider both descriptive discriminant analysis (DDA) and predictive discriminant analysis (PDA); these are two very different applications (see Huberty, 1994; Thompson, 1995a). It is suggested that stepwise methods should not be used in multivariate statistics.

Although discriminant analysis (DA) is often referred to as a single technique, DA actually encompasses two discrete methods, each having greatly different functions. Huberty (1994), who quite literally "wrote the book" on Discriminant Analysis, describes the two methods with distinctly different purposes--one pertaining to *predictive discriminant analysis* (PDA), and one pertaining to *descriptive discriminant analysis* (DDA).

> Briefly, descriptive discriminant analysis (DDA) pertains to the problem of explaining and interpreting effects found via multivariate analysis of variance (MANOVA). With this analysis (DDA), the response variables play the role of criterion variables. Predictive discriminant analysis (PDA), on the other hand, pertains to the problem of group membership prediction or classification. In PDA, the response variables play the role of predictor variables. (Huberty, 1989, p. 54)

Huberty goes on to state that there are actually very few commonalities between PDA and DDA: "About the only characteristics common across all meanings of discriminant analysis are: (a) there are multiple response variables and (b) there are multiple groups of objects or subjects. Thus, in any discriminant analysis context there are two sets of variables; one set consists of a collection of response variables, the other set consists of one or more grouping or nominally scaled variables" (p. 158). However, the roles of the predictor and criterion variables are reversed across these two very distinct applications.

A brief description of PDA and DDA and their purposes is presented below for the purpose of emphasizing distinctions between them, and to facilitate discussion of the use (and misuse) of stepwise methods in both PDA and DDA. However, this paper is not intended as a tutorial on discriminant analysis, and a basic familiarity with these methods

is assumed. More complete descriptions and discussions of PDA and DDA techniques can be found elsewhere (Huberty, 1984, 1989, 1994; Huberty & Wisenbaker, 1992; Klecka, 1980).

Predictive Discriminant Analysis (PDA)

PDA comprised the primary original use of discriminant analysis (Huberty & Wisenbaker, 1992), and focuses on classification of objects or data sets into one of two or more well-defined criterion groups. PDA can be used for two purposes: (a) to determine effective estimates that result in highest group classification, commonly called the hit rate; and (b) to determine whether the hit rate obtained with the selected estimators is greater than would be obtained by chance (Huberty, 1984).

The primary statistic used in PDA is Linear Classification Functions, or LCFs, although Quadratic Classification Functions (QCFs) are used in the case of unequal covariance matrices. (Young (1993) provides an explanation of the difference between "linear" [uses a pooled covariance matrix] and "quadratic" [uses a separate covariance matrix for each group] functions.) Each case receives LCF scores equal to the number of groups (*k*), and is then predicted to be part of the group for which the case received the highest LCF score (Huberty, 1984). These LCF predictions for all cases are then tabulated and presented in a classification table, which is used to characterize PDA "hit rates" (Huberty & Barton, 1989, p. 161). A *hit* results when a case is correctly assigned to its own group.

The hit rate obtained through PDA is then compared to the hit rate that would be obtained by chance as a means of evaluating the effectiveness of the set of predictor variables used. Further analyses should be conducted with subsets of predictor variables,

called *leave-one-out* analyses, to see if the hit rate can be maintained, or even improved, by using a smaller set of predictor variables (Huberty & Wisenbaker, 1992). (Stepwise methods are often used to select a smaller set of predictor variables, but this is an erroneous use, as will be discussed later.) An anomaly of PDA is that additional variables can actually hurt the hit rate, which places PDA somewhat outside of the General Linear Model (GLM). This aspect of PDA is beyond the parameters of this paper, but is explained elsewhere by Thompson (1995a, 1998).

The most concise subset of variables that can maintain a comparably effective hit rate can then be considered a good subset *for that particular sample*. As Huberty (1989) pointed out,

> It is somewhat common to find in the applied literature results of a PDA based on *internal* classification, that is, where the classification rule employed was built on the very cases that were used in obtaining the classification table... A much better practice in most situations is to employ an *external* analysis, where the cases used to generate the classification table are different from those used to build the classification rule. (p. 161)

Methods used to test generalizability of a classification rule include cross-validation methods or re-sampling techniques such as bootstrap or jackknife methods. For further discussion of these methods, see Thompson (1989, 1999) and Huberty (1989).

Descriptive Discriminant Analysis (DDA)

DDA, by comparison, focuses on describing, explaining, or interpreting (rather than predicting) the effects of a grouping variable(s) on a set of criterion variables. DDA is most often used to describe and interpret overall or group effects found when a multivariate analysis of variance (MANOVA) is conducted (Huberty & Barton, 1989; Huberty & Wisenbaker, 1992). DDA can be used for three purposes: (a) to select a subset

of response variables that yields nearly the same grouping effects as the original larger set; (b) to order the reduced set of variables by their strength of contribution to the grouping effects; and (c) to explain the underlying structure of the grouping effects (Huberty, 1984).

Huberty (1984; Huberty & Barton, 1989) cited two basic sets of statistics with regard to DDA. These are (a) linear discriminant functions (LDFs) and structure coefficients and (b) Wilk's Lambda. Wilk's lambda is "the typical index used to assess the goodness of a variable subset" (Huberty, 1989, p. 45) and is inversely analogous to the squared multiple correlation coefficient (R^2) in the regression case. An LDF (note: *not* the same as a LCF used in PDA) is a linear composite representing multiple outcome variables (Huberty & Barton, 1989); that is, LDF (or QDF) scores are obtained for each case on each LDF by applying the LDF additive and multiplicative scores to the scores of each case on the measured outcome variables. Huberty (1989, p. 163) argued that "These composites [LDFs] define scores on latent constructs that underlie resultant MANOVA effects, and an effort should be made to interpret these latent constructs since they are the actual focus of the analysis." LDFs are obtained by an eigenanalysis, which is discussed in Huberty (1994). In multi-way layouts, a separate set of LDF's will be obtained for each main and interaction effect (Huberty, 1984). LDF's are often referred to simply as "functions," although with the proliferation of terms including the word "function" in them, this can lead to confusion. The number of LDF's for each effect is limited to the number of response groups minus one *(k-1)*, as opposed to *(k)* in PDA (Huberty, 1984; Thompson, 1998).

The group "centroids" (i.e., mean LDF or QDF scores on each function) for each response variable can be plotted graphically on a territorial map of the LDF's to better see the relationships between each variable with regard to the LDF's. Succinctly stated, an LDF is a synthetic variable representing an underlying trait or characteristic that represents aspects of one or more variables. By studying the location of each variable's group centroid in relation to the LDF's, a researcher can get an idea as to which specific variables contribute to each LDF. The Discriminant function scores, which are analogous to Yhat scores in multiple regression (Thompson, 1998), can be consulted as a means to assist in identifying either "fenceriders" or "outliers". There is a big HOWEVER, however-- a researcher cannot examine *only* the LDFs to understand the complex relationship of response variables to the functions.

Thompson (1992; Thompson & Borrello, 1985) stressed emphatically the importance of consulting structure coefficients along with beta weights as aids to interpretation in regression applications. In DDA, structure coefficients are analagous to structure *r's* in regression, and are therefore equally vital to consult along with LDF's for substantive interpretation of response variables. Response variables may be collinear with regard to their LDF values, so an important contributing variable may be overlooked. Structure coefficients are simple bivariate correlations, so collinearity is not a problem. Simply put, a structure coefficient reveals "how closely a variable and a function are related" (Klecka, 1980, p. 31).

Huberty (1994) emphasized that "construct definition [i.e., LDF or QDF coefficients] and structure dimension [i.e., structure coefficients, and <u>not</u> hit rates] constitute the <u>focus</u> of a descriptive discriminant analysis" (p. 206, emphasis added) . He

also observed that, given the general linear model and the identities of techniques across methods (Thompson, 1998) :

> If a researcher is convinced that the use of structure *r*'s makes sense in, say, a canonical correlation context, he or she would also advocate the use of structure *r*'s in the contexts of multiple correlation, common factor analysis, and descriptive discriminant analysis. (p. 263)

From the previous discussion, it seems sufficiently obvious that PDA and DDA are two *very* different applications, with different purposes, different roles of variables, and that completely different statistics are consulted in result interpretation (see also Huberty, 1984, 1989, 1992, 1994; Thompson, 1995a, 1995b, 1998, 1999). Table 1 summarizes the contrasts between PDA and DDA techniques.

Stepwise Methods

The idea of reducing the number of response variables used to explain a particular effect-- often referred to as variable selection--reflects the idea of parsimony, which is a guiding principle of science (Babbie, 1990). A second concept, variable ordering, refers to the ranking of variables according to their strength of effect. While these concepts are attractive, some researchers mistakenly use so-called stepwise methods as a Holy Grail to guide decisions regarding the selection, organization, and categorization of their variables in both PDA and DDA. As Huberty (1989, p. 45) pointed out, "Stepwise analyses have basically been used for three purposes: (1) selection or deletion of variables; (2) assessing relative variable importance; or (3) both variable selection and variable ordering. Note that these are uses often made of stepwise analyses, not uses that necessarily *should* be made". Thompson (1995b, p. 526) concurred, noting that "stepwise methods are not usually useful for either purpose."

Table 1
Comparison of Predictive (PDA) and Descriptive (DDA) Methods of Discriminant Analysis

	PDA	DDA
Purpose	Classification or prediction of group membership	Interpretation/Explanation of group separation or group differences
Role of Response Variables	Predictor Variables	Criterion Variables
Role of Grouping Variables	Criterion Variables	Predictor Variables
Questions Answered	(a) What are good estimates of separate-group and total-group percents of correct classifications (i.e., hit rates)? (b) Are the observed hit rates better than those expected by chance? (Huberty, 1984, p. 157)	(a) Is there a subset of the original set of response variables that yields nearly the same effects as the original set? (b) What is a reasonable ordering of the retained set of variables in terms of their relative contribution to the resulting effects? (c) What is a reasonable substantive interpretation of the structure underlying the resulting effects? (Huberty, 1984, p. 156)
Primary Statistics Used	Linear Classification Functions (LCFs), Classification Table	Linear Discriminant Functions (LDFs), Wilks Lambda, Structure Coefficients
Number of Values	# LCFs = *(k)*	# LDFs = *(k-1)*
Focus of Analysis	Hit rate	LDF Coefficients, Wilks Lambda
Other		Used primarily following a MANOVA to explain resulting effects

The essence of stepwise methods is that a series of models--or LCF (or QCF) or LDF (or QDF) equations, in the case of discriminant analysis--are developed, with response variables entered one step at a time (Huberty, 1989; Thompson, 1989). In the first step, the single best response variable is selected. Various statistics can be used for

discriminating the "best" response variable (Klecka, 1980), although lambda is most often used, with the variable that reduces lambda the most being selected as the first "best" response variable. In subsequent steps the "next best" response variable is selected by consulting the remaining variables and selecting the one that reduces lambda to the greatest extent when added to the set already selected.

In true stepwise methods, variables entered are subsequently considered for removal at each step. The most commonly used form of stepwise method, however, is actually a forward selection procedure, which does not consider variables for removal once entered. Huberty (1989, p. 44) stated, "Although many researchers claim to have used a stepwise procedure, what they in fact used was simply a forward selection procedure". Backward selection procedures are also available, but are seldom used.

Numerous cautions, indictments, warnings, and admonitions advising against the use of stepwise methods have been recorded in the literature (Huberty, 1984, 1989, 1994; Huberty & Barton, 1989; Huberty & Wisenbaker, 1992; Snyder, 1991; Thompson, 1995a, 1995b, 1989, 1998, 1999). Huberty offered a more specific warning pertaining to the use of stepwise methods with PDA:

> Another shortcoming [of stepwise procedures] pertains to the use of available computer packages to perform stepwise analysis with PDA. Criteria used in stepwise programs in all three packages [SPSS, SAS, and BMDP] pertain to DDA, if anything, but *not* to PDA. The variable selection and ordering results from such programs have sometimes been illegitimately used for PDA purposes.
> (1992, p. 177)

Thompson (1989, 1995a, 1995b, 1998, 1999) discussed three primary problems with stepwise methods. The remainder of this paper discusses these problems in relation to discriminant analysis, and provides analogies of these problems in accessible non-

mathematical terms in an attempt to promote understanding of the egregious errors that can occur with these methods and to discourage their use. Primary emphasis on the use of stepwise methods with DDA techniques rather than PDA, since its use with PDA is already shown to be inappropriate.

Wrong Degrees of Freedom

As Thompson (1995a) explained, "Degrees of freedom in statistical analyses reflect the number of unique pieces of information present for a given research situation. These degrees of freedom constrain the number of inquiries we may direct at our data and are the currency we spend in analysis" (p.526). The idea of degrees of freedom as currency is an attractive one, as the consequences of spending unwisely or incorrectly are easily understood. A major problem with computer packages implementing discriminant function analysis with stepwise methods is that they use the wrong degrees of freedom in their statistical tests. This is regrettable because, as Thompson (p. 526) noted, "The use of incorrect degrees of freedom in practice often has dire consequences regarding the accuracy of our inferences." To keep with the currency analogy, the buyer is getting a deal that sounds too good to be true, and indeed is!

As stated earlier, when selecting the first "best" response variable, the lambdas of *all* of the variables are consulted to make the selection. Yet, when the variable resulting in the lowest lambda is selected for entry into the discriminant function equation, the computer package only "charges" for the use of one response variable. At each subsequent step all of the remaining variables are consulted, with the variable that yields the greatest additional reduction in lambda selected for entry into the equation. A single degree of freedom is added at each step, reflecting the actual single variable that is added

rather than the total number of variables that were consulted for entry. This is simply incorrect!

In an extensive discussion of this topic, Thompson (1998) compared this to a cafeteria line, where all of the food items are variables. A customer (researcher) cannot just sample food (variables) at random and then pay (with degrees of freedom in the numerator) for only what (s)he liked! Of course, all items of food sampled, or variables consulted, must be paid for with the appropriate degrees of freedom (coin) spent on each item. As he stated,

> The statistical significance tests take into account both the number of coins we've chosen to spend [degrees of freedom between, in the numerator] and the number we have chosen to reserve [degrees of freedom within, in the denominator]. The most rigorous tests occur when we spend few degrees of freedom [in the numerator] and reserve many [in the denominator]. (p. 17)

In other words, the rigorous researcher is rewarded for keeping a large savings account of degrees of freedom in their denominators!

Thompson (1998) further illustrated this with an example of a study involving two steps of stepwise discriminant analysis, with k=3 groups and n=120 people. In each step, a single degree of freedom is added to the numerator for a total of 2 degrees of freedom in the numerator, making the test seemingly much more rigorous than it really is. In the correct version, 100 degrees of freedom is correctly added to the numerator--a tremendous difference from 2! The results are obviously very different; in the stepwise version (with the unfairly easy test) statistical significance is found with F=13.64322 and p calc= .0000945. In the correct version, the results are not statistically significant, with F= 0.31991 and p calc= 1.00000. This is comparable to having all the correct answers

prior to taking an exam. The resulting grade may be good, but it is not a rigorous testing of the student's knowledge and the grade received doesn't reflect the student's actual knowledge. To put it in a more common vernacular, it's like cheating!

Thus in stepwise methods, the degrees of freedom are systematically biased in favor of yielding spuriously statistically significant results, with the inaccuracy compounded through errors in both the numerator and denominator. As Thompson (1995b, p. 527) explained,"...the use of the incorrect degrees of freedom can (a) radically inflate MS explained, (b) radically deflate MS unexplained, and consequently (c) radically inflate F calculated.... This statistical welfare system may cause us to radically overestimate the atypicality of our results (i.e., create an artificially small p calculated)." In the related regression case, Cliff (1987, p. 185) surmised that "most computer programs for [stepwise] multiple regression are positively satanic in their temptations toward Type I errors." The same statement can certainly be applied in the discriminant analysis case. Unfortunately, most students using packaged computer programs probably do not realize the error and presume the program output to be correct.

<u>Does Not Identify the Best Predictor Set of Size "q"</u>

It is widely, albeit incorrectly, presumed that the steps in a stepwise analysis select the best subset of response variables. For example, if three steps are conducted, then the three response variables selected must be the best set of three. Not true! The only correct aspect here is that the first response variable selected is the best single response variable *of the ones in the study*. The second response variable selected is the one that minimizes lambda the most *in the presence of the first selected response variable.* Likewise, the third response variable selected is the one that again minimizes lambda to

the greatest extent *in the presence of the first two response variables.* This is completely different from selecting a 'best subset' of three response variables, which in fact, stepwise selection methods do not do (Huberty, 1989; Snyder, 1991; Thompson, 1995a, 1995b, 1998). In fact, SPSS version 10.0 (1999), a commonly used statistical computer package, cautions users in its manual to "...be aware that none of these [stepwise] procedures is guaranteed to provide the *best* subset in an absolute sense" (p. 216).

The issue of the best subset selection can be considered in the context of the Olympic games. In an event, the best athlete (best response variable) is selected as the gold medal winner, with the silver award going to the next best contender (second best response variable) *in the presence of the gold medal winner*, and the bronze medal is awarded to the next best contender (third best response variable) *in the presence of the first two.* The awards represent a ranking, *not a team*, of athletes who performed the best in that particular contest on that particular day, and thus present a situation-specific context.

Few would make the mistake of considering the medal winners collectively as representing the best <u>team</u> of athletes, so why interpret the selected and ranked predictor variables in a stepwise discriminant analysis as the best <u>subset</u> of predictors? Further, if the gold medal winner, for example, was very narcissistic and liked to hog the ball at every opportunity, the best three-member team might well consist of the silver and bronze medal winners and yet another player. The best three-member team in this instance might <u>not</u> even include the gold medal player, because even though this person played best alone, this person might actually make the team play less well together.

A spectator who assumes the medal winners are truly the best three-person team in the world for that particular event might very well be wrong. With regards to variable selection, Huberty (1989, p. 47) likewise admonished, "If one or two of the variables already entered would be changed, then the third variable entered may also be different. This dependence or conditionality truly makes variable importance as determined by stepwise analyses *very* questionable."

In fact, the best variable set of a given size (a) may yield considerably higher effect sizes than the stepwise variable set of the same size and (b) may even include <u>none</u> of the predictors selected by the stepwise algorithm (Thompson, 1995b)! As regards variable subset selection, alternative stategies advocated elsewhere (Huberty, 1994; Huberty and Wisenbaker, 1992; Thompson, 1998) "all possible subsets" approaches, so that the researcher may examine the results of all possible multiple combinations of variables and select a desirable subset that meets his or her conditions.

<u>Nonreplicability of Results</u>

The final problem with using stepwise methods with discriminant analysis involves the generalizability of stepwise results, or the lack thereof, and relates to the variable selection problem discussed previously. The variables selected for inclusion in a stepwise analysis are chosen in a <u>situation-specific context</u> (Thompson, 1998) . That is, adding one additional response variable or deleting one single response variable from the analysis may entirely change not only the order of variable selection, but even what variables are selected. Because we are never exactly sure that we have <u>all</u> the correct variables and <u>only</u> the correct variables (i.e., part of a "correctly-specified model"), the context-specificity of stepwise results is very troubling! Furthermore, in many research

situations, several variables may present minute differences with regards to their effect on lambda. For example, the decrease in lambda between two variables may differ by only .001 (one thousandth of a percent!) yet the smaller lambda is chosen over the larger one as a "better" response variable. Thus, stepwise methods capitalize outrageously on even small amounts of sampling error, thus yielding results that will not generalize beyond the sample (Snyder, 1991).

This reliance on small differences between variables can be illustrated by considering a childhood game called "Gossip." In this game children sit in a circle, and a silly phrase is whispered to one child. The first child whispers the phrase as quietly as possible to the next child, and the phrase is passed around the circle in this manner, with the final child announcing the phrase that they heard for comparison with the original. No one is allowed to repeat the phrase to his or her neighbor more than once. If a child doesn't understand a word, or misjudges the phrase, the mistake is passed on, with the result that each error is compounded. By the time the phrase reaches the end of the circle, it usually bears little resemblance to the original one.

Similarly in stepwise methods, small amounts of sampling or measurement error in response variables can result in erroneous variable selection and ordering through over-reliance on small differences in lambda. Thompson has repeatedly stressed (1989, 1995a, 1995b, 1998) that small amounts of change in lambda can easily be accounted for through error, yet these minute differences are allowed to guide the process of variable selection and ordering in stepwise procedures! The final predictor set selected is fitted *only* to the data at hand, and not to the research problem in general, and produces results that are unlikely to generalize. One would not place any amount of reliance on the final

outcome phrase of a kindergardener's game of gossip; likewise, the replicability of stepwise results should be regarded with similar skepticism.

Summary

The possibility of a method that will select and order predictor variables (without any thought needed on the part of the researcher!) can, unfortunately, make stepwise methods seem very attractive to unsuspecting researchers. Kerlinger (1986) provided sage advice for students considering using stepwise methods when he argues that "the research problem and the theory behind the problem [and not stepwise methods] should determine the order of entry of variables" (p. 545) . Stepwise methods pose serious problems, and are used incorrectly by most researchers. In truth, there is no Holy Grail for variable selection and ordering. Careful consideration of the variables and thoughtful examination of the data must be invoked with regard to the reality of the research problem being considered.

REFERENCES

Babbie, E. (1990) . Survey Research Methods (2nd ed.). Belmont, CA: Wadsworth, Inc.

Cliff, N. (1987) . Analyzing multivariate data. San Diego: Harcourt Brace Jovanovich.

Huberty, C.J. (1984) . Issues in the Use and Interpretation of Discriminant Analysis. Psychological Bulletin, 95(1), 156-171.

Huberty, C.J. (1989) . Problems with stepwise methods--better alternatives. In B. Thompson (Ed.), Advances in social science methodology (Vol. 1, pp. 43-70). Greenwich, CT: JAI Press.

Huberty, C.J. (1994) . Applied discriminant analysis. New York: Wiley and Sons.

Huberty, C.J., & Barton, R.M. (1989) . Methods, Plainly Speaking: An introduction to Discriminant Analysis. Measurement and Evaluation in Counseling and Development. (22), 158-168).

Huberty, C.J. & Morris, J.D. (1989) . Multivariate analysis versus multiple univariate analyses. Psychological Bulletin, 105, (302-308).

Huberty, C.J., & Wisenbaker, J.M. (1992) . Discriminant Analysis: Potential improvements in typical practice. In B. Thompson (Ed.), (1992). Advances in educational research: Substantive findings, methodological developments (Vol. 2, pp. 169-208). Greenwich, CT: JAI Press.

Kerlinger, F.N. (1986) . *Foundations of behavioral research* (3rd ed.) . New York: Holt, Rinehart and Winston.

Klecka, W.R. (1980) . Discriminant Analysis. Newbury Park, CA: Sage.

Snyder, P. (1991) . Three reasons why stepwise regression methods should not be used by researchers. In B. Thompson (Ed.), (1991). Advances in educational research: Substantive findings, methodological developments (Vol. 1, pp. 99-105). Greenwich, CT: JAI Press.

Statistical Package for the Social Sciences®. (1999) . SPSS® Base 10.0 Applications Guide. Chicago, IL: SPSS Inc.

Thompson, B. (1989) . Why won't stepwise methods die?. Measurement and Evaluation in Counseling and Development, 21(4), 146-148.

Thompson, B. (1992, April) . Interpreting regression results: beta weights and structure coefficients are both important. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 344 897).

Thompson, B. (1995a) . Review of Applied discriminant analysis by C.J Huberty. Educational and Psychological Measurement, 55, 340-350.

Thompson, B. (1995b). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. Educational and Psychological Measurement, 55, 525-534.

Thompson, B. (1998, April) . Five methodology errors in educational research: the pantheon of statistical significance and other faux pas. **Invited address** presented at the annual meeting of the American Educational Research Association, San Diego. (ERIC Document Reproduction Service No. ED 419 023)

Thompson, B. (1999, April) . Common methodology mistakes in educational research, revisited, along with a primer on both effect sizes and the bootstrap. **Invited address**

presented at the annual meeting of the American Educational Research Association, Montreal. (ERIC Document Reproduction Service No. ED 429 110)

Thompson, B. & Borello, G.M. (1985) . The importance of structure coefficients in regression research. Educational and Psychological Measurement, 45, 203-209 .

Young, B. (1993, November) . Quadratic versus linear rules in predictive discriminant analysis. Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 364 686)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com